Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 4, 2023
A Regional Conference of the Society for Industrial and Systems Engineering

Modeling Assessments of Special Operations Aviation Candidates

Christopher Letarte, John Lucero, and Si Park

Department of Mathmatical Sciences
United States Military Academy
West Point, NY 10996

Corresponding author's Email: christopher.letarte@westpoint.edu

Author Note: Sincerest gratitude to the United States Army Special Operations Aviation Command (USASOAC) and all of the personnel there who assisted with our project. The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract:

The Special Operations Aviation Regiment (SOAR) is renowned for its involvement in some of the most important military missions across the globe. Given recent recruiting shortfalls in the U.S. Army, the Special Operations Aviation Training Battalion (SOATB) is entrusted with the increasingly difficult task of selecting the best candidates for SOAR. Our goal was to improve training efficiency within one of SOATB's training programs, Enlisted Green Platoon (EGP), by determining the most critical predictors of success among candidates. Using machine learning models, we found that a candidate's psychometric score was highly significant in predicting graduation from the program.

Keywords: Predictors, Psychometric Variables, Logistic Regression, Random Forest

1. Introduction

The United States Army is currently facing a daunting task: recruiting tomorrow's soldiers to meet the Army's in- creasing demands within a highly competitive and complex operating environment. Recent recruiting challenges have greatly impacted the special forces community. While many special forces candidates enter the U.S. Army directly through a contract, others come from different branches of the U.S. Army. As the pool of recruits eligible to join the special forces community dwindles, the quality of candidates will likewise decline (Winkie, 2022).

The mission of the USASOAC is to equip, train, resource, and force-generate Army Special Operations Aviation forces. To accomplish this demanding mission, SOATB, which falls under SOAR, conducts specialized individual training and provides education in order to produce crew members and support personnel with basic and advanced qualifications. This high level of training and qualification is carefully planned and executed to identify high-achieving soldiers through the selection process known as EGP. This training program is six weeks long and serves as the regiment's initial entry training and testing program, where candidates are introduced to SOAR's creed, which emphasizes a 'Never Quit' attitude based on grit (Tracy & Esposito, 2019). To minimize costs and increase the efficiency of selection processes, it is critical to improve the regiment's ability to identify quality candidates with an increased likeliness of success. Our study utilizes machine learning techniques to analyze recent EGP candidates and identify the attributes that best predict selection as a future special operations aviation soldier.

EGP candidates start the program in the assessment platoon, where each candidate undergoes a series of batteries and examinations. If a soldier is deemed qualified in all administrative categories, they will be given a "Go for Train." Next, the candidate will enter one of four training modules managed by their respective cadre committees. Each module has an academic and performance-related assessment. Throughout EGP, there are additional physical events to assess endurance and strength. Once individuals graduate from selection, they progress towards follow-on technical training or report to SOAR (Tracy & Esposito, 2019).

1.1. Literature Review

A previous study used both logistic regression and random forest models to analyze the effect of physical and psycho- logical factors on the graduation of special operations candidates from the Ranger Assessment and Selection Program (RASP). This study utilized a detailed breakdown of psychological factors (~ 150 psychometric variables) for each individual student. They determined the most critical indicators of success in RASP were physical fitness scores, especially those from a candidate's most recent Army Physical Fitness Test (APFT). The most important psychometric factor was ranked seventh in importance in the model. The results of the analysis highly emphasized the importance of being physically fit over the several other variables being considered. Both machine learning models were able to successfully predict the outcome of the candidates about 80% of the time (Vinnedge, Schwartz, Baller, & Dykhuis, 2022).

Another study used a composite psychological score to assess the psychological hardiness of candidates admitted to West Point. The study performed well at predicting leader adaptability among officers who graduated from the academy. This research presented a potential association between psychological hardiness and success as military officers. These findings offer useful inspiration for further research in the applicability of psychological metrics in predicting success in the military (Bartone, Kelly, & Matthews, 2013).

2. Methodology

2.1. Data Collection

The data was collected from May 2020 to September 2022 on all soldiers who attempted EGP. The initial data set contained 2145 observations of 115 variables. This data included administrative data, EGP training scores, physical fitness scores, and a single psychometric score. We sourced this data from multiple internal databases maintained by SOATB's database manager. To understand the data collected, we visited the SOATB and conversed with leaders of the different training committees and the individuals in charge of the data sets. This gave us a holistic view of the training and the reasoning for conducting this analysis. The response variable is a binary categorical variable that indicates whether the individual graduated from EGP. Upon discussion with SOATB leadership, we initially determined that the most important variables were the General Technical (GT) score, summary psychometric score, and each of the four EGP mod scores; we later identified sex and MOS as important variables to control for.

An additional set of variables that were collected but not used in analysis were grit, integrity, tact, and trainability (GITT) variables. During each EGP module, candidates were evaluated on these qualities. These would have served as excellent predictor variables but had two issues. The first issue is that they were recently implemented, and the second is that the variables appeared not to represent what they were designed to. Instructors were forced, due to staffing, to score everyone as average except for in extreme cases. This is a useful tool for the instructors but does not objectively measure the GITT of the candidates.

2.2. Data Cleaning

The initial data set included many missing and incorrectly entered observations. The first step was identifying which variables were of interest based on what our sponsor wanted us to analyze. Initially, our variables of interest included: psycho- metric score, GT score, Military Occupational Specialty (MOS), module score, and their impact on graduation.

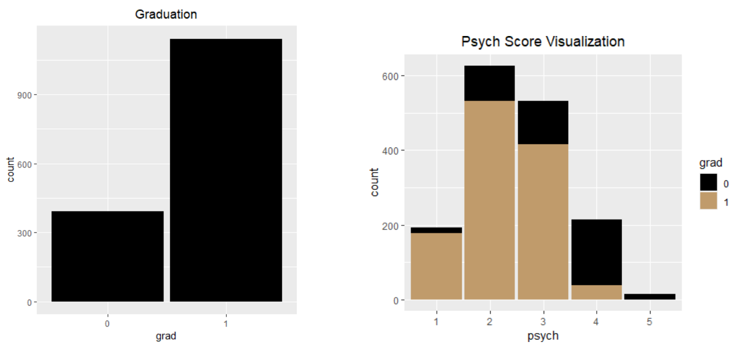
The second step in cleaning the dataset was to remove any repeat observations. Since it was possible for a candidate to fail EGP and be allowed to re-enter later, some individuals had several entries in the dataset. By comparing administrative data such as birth date, sex, MOS, and initial service date, it was possible to identify and eliminate individuals who appeared multiple times. The most recent observation was kept for each individual, whether that was a pass or a fail. This reduced the dataset to 1820 observations but eliminated much of the dependence observed in our dataset. Next, we removed any observations that contained null values leaving 1534 observations.

A close analysis of module variables revealed some issues. The module scores were collected in sequence with the EGP experience, and when candidates fail, this leaves incomplete observations for follow-on training. The final module in EGP was left as a null value in each case where a candidate failed. The individual module scores were excluded to eliminate the covariance, and the total score (sum of the individual module scores), was used instead.

The response variable, whether a candidate ultimately ended up graduating from EGP, had a pass rate of 74% based on individuals who started EGP. This rate does not include individuals who recycled or were not allowed to start EGP.

Lastly, the quantitative psychometric variable was transformed into an ordered factor, with a 1 (the highest score) being the base and each subsequent score $(2 - 5)$ as a separate factor. The MOS of each candidate was turned into a binary factor representing whether or not they were in a high-density MOS, meaning that at least 60 candidates had been in that

MOS since class E20-06. The ACFT variable was also transformed into a factor with four levels: 1X NO GO (meaning they failed and then passed), 2X NO GO (meaning they failed twice before passing), a NO GO (meaning they failed entirely), and a GO meaning that they passed on the first attempt.



(a) Response Variable (b) Psych Score

Figure 1: Most Significant Categorical Variables

2.3. Data Processing

The Numerical variables were all converted to float values and each categorical variable was turned into a factor. Next, the data was split into train and test sets, with 20% of the data being used in validation. The test set was used to calculate all performance metrics and evaluate variable importance.

2.4. Analysis Methods Used

The methods used in the predictive analysis were a logistic regression model and a random forest model.

2.4.1. Logistic Regression

$$ln\left(\frac{P}{1-P}\right) = \beta_0 + \beta_1 \times x_1 + \beta_2 \times x_2 + \begin{cases} \beta_{3_1} \times Psych2 \\ \beta_{3_2} \times Psych3 \end{cases} \tag{1}$$

A logistic regression model is a statistical analysis used for predictive analysis and modeling. In a logistical modeling approach, the dependent variable must be categorical, either binary or multinomial. Logistic regression models the probability of a positive result by calculating the optimal coefficients for each explanatory variable.

These models are better suited to compute quantitative variables with a disadvantage in computing categorical variables. Since each variable is fit, quantitative variables have coefficients that allow a continuous output, while categorical variables are given fixed levels. (Thanda, 2022)

2.4.2. Random Forest

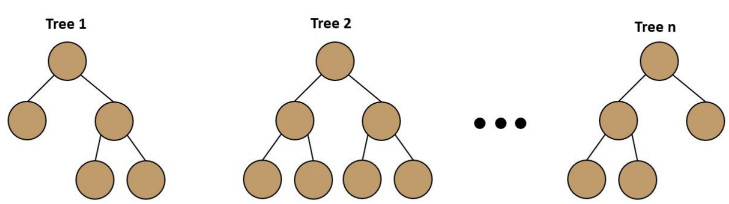


Figure 2: Random Forest Diagram

ISBN: 97819384964-4-8

A random forest model utilizes feature randomness to generate a set of trees with unique permutations of variables. This randomness leads to a model which produces different, sometimes substantially better results. Each set of trees has a maximum depth, with the leaves of the trees being the class that the individual decision tree determines to be best fit for an observation. The standard number of trees is five hundred, and each of these trees is optimized to classify given a unique set of variables and order in which to assess them (Yiu, 2019).

Random forest models are better equipped to handle data that has multi-level categorical variables, with a disadvantage in handling numerical data. Each node in a decision tree is split on features: categorical variables on their unique levels and numerical data on calculated decision points.

3. Results

3.1. Comparison of Models

We tested several iterations of logistic regression and random forest models. The metrics used in measuring success were the accuracy of the model and the Receiver Operating Characteristics - Area Under the Curve score (ROC-AUC score). Accuracy is the simpler of the two metrics. Accuracy is calculated by dividing the correct number of predictions by the total number. Accuracy can be calculated from the confusion matrix generated in each model by looking at adding together the True Positive assessments (TP) and True Negative Assessments (TN) and dividing them by the sum of the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) Assessments. Accuracy is a logical assessment yet it fails to measure how well the model is able to differentiate between classes (Ragan, 2018).

$$Accuracy = \frac{TN + TP}{TN + TP + FN + FP} = \frac{Correct\ Predictions}{Total\ Predictions} \tag{2}$$

The ROC-AUC metric is a measurement of performance for how much a model is capable of distinguishing between classes. A ROC curve plots the true positive rate of a model against its false positive rate. A model's true positive rate or sensitivity is a measure of how good a model is at correctly identifying an individual who will pass. The false positive rate is a measure of how often a model incorrectly labels a negative outcome as a positive outcome. By plotting the two rates against one another, it allows for a visual representation of how good a model is at separating positive and negative results. The ROC-AUC score serves as a numerical representation of the ROC curve and serves as a metric that represents the ability of a model to correctly distinguish positive and negative outcomes (Narkhede, 2018).

$$Sensitivity/True\ Positive\ Rate = \frac{TP}{TP + FN} = \frac{Correct\ Positive\ Predictions}{Total\ Positive\ Values} \tag{3}$$

$$False\ Positive\ Rate = 1 - Specificity = 1 - \frac{TN}{TN + FP} = 1 - \frac{Correct\ Negative\ Predictions}{Total\ Negative\ Values} \tag{4}$$

Both logistic regression and random forest performed well. The logistic regression model had an accuracy of 0.9498 and a ROC-AUC score of 0.982. The logistic model can correctly predict 94.98% of the candidate's outcomes. The random forest model had an accuracy of 0.9635 and a ROC-AUC of 0.981. The accuracy of the random forest model is slightly better than the logistic regression model, correctly predicting 96.35% of the candidate's outcomes. Although logistic regression performed slightly worse, both models performed well with excellent accuracy and ROC-AUC scores (Narkhede, 2018).

Table 1: Model Summary Statistics

	Accuracy	**ROC-AUC**
Logistic Regression	0.9498	0.982
Random Forest	0.9635	0.981

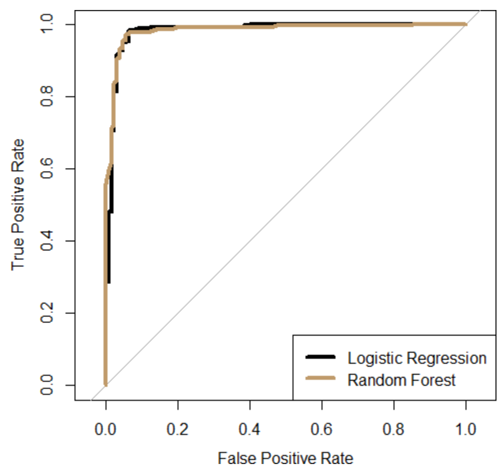


Figure 3: Receiver Operator Curve for Linear Regression and Random Forest Models

3.2. Analysis of Best Model

After comparing our two models, we found the random forest model performed better at predicting the correct outcome. However, the logistic regression model was better able to attribute the importance of the variables based on the amount of variance explained. For analysis, the variables were ranked on their importance (the amount of total variance they explained) (Carnell, 2021).

Table 2: Model Summary Statistics

Variable	**Importance**
Total Score	50.9%
Psych	28.5%
ACFT	19.8%
Sex	0.50%
MOS Density	0.30%
GT Score	0.01%

The variable that explained the most variance was the candidate’s Total Score. The total score is the sum of the candidate’s individual module scores. Since an individual needs to pass each of the modules to pass EGP, it follows that scoring well in the modules is predictive of success. The next most impactful variable is the psychometric variable, which is the best proxy for grit contained within the data. The last majorly impactful variable is whether the individual passed the fitness test (ACFT/APFT) at the beginning of EGP. The final three variables: Sex, MOS Density, and GT Score each explained negligible amounts of variability but were identified as important control variables.

4. Conclusion

4.1. Analysis of Results

The best model shows that the most important factor is the psychometric score. The non-numerical physical variable demonstrates the institutional importance of physical fitness as more of a minimum standard that is required to train effectively instead of a factor that is essential in a candidate’s ability to complete the module. While the total score explains the most variance, it is not useful in the analysis of EGP.

ISBN: 97819384964-4-8

4.2. Implications and Future Research

This model is not designed as a tool with which to eliminate candidates who are attending EGP. Instead, this model serves as a tool that SOAR can utilize to identify the most important factors in succeeding in EGP. While SOAR is actively searching for soldiers with grit who possess the ability to perform under stress, this may prove to be a significant endeavor. SOAR introduced GITT as a metric to identify grit within candidates, but insufficient data collection prevented us from incorporating it into our analysis. However, it may be possible to use other collected information to conduct a more detailed analysis to identify other attributes in the future. Bartone created a composite psychological score to determine the psychological hardness, and the current psychometric score computed by regimental psychologists serves a similar purpose.

The differences in individual modules lend themselves to future research. Since each instructional module covers a different subject but shares a similar structure, there may be systematic issues that are decreasing the effectiveness of EGP. Future analysis that could prove useful to SOAR would be to analyze each individual module and identify the most impactful events in each and identify the events that are least important. Increasing the efficiency of EGP would improve the cost-effectiveness and enable more effective recruitment of talent. Additionally, identifying issues in individual models could lead to the identification and improvement of systemic issues in EGP.

Another opportunity for further research is the possibility of long-term success, which could be measured by time with the unit or how well they performed in evaluations. With the issues of both recruitment and retention in the SOAR and the Army at large, looking at the impact of a candidate's initial assessment and performance during their time in EGP on their overall success in the unit could be quite useful. It would improve the Army's ability to recruit the ideal candidates best suited for a career in the SOAR.

References

Bartone, P. T., Kelly, D. R., & Matthews, M. D. (2013). Psychological hardiness predicts adaptability in military leaders: A prospective study. *International Journal of Selection and Assessment*, *21*(2), 200-210.

Beal, S. A. (2010). *The roles of perseverance, cognitive ability, and physical fitness in us army special forces assessment and selection* (Tech. Rep.). Army Research Institute for the Behavior and Social Sciences.

Carnell, R. (2021). *Tornado r package documentation.* Retrieved from https://github.com/bertcarnell

Kleinbaum, D. G., Dietz, K., Gail, M., Klein, M., & Klein, M. (2002). *Logistic regression*. Springer.

Narkhede, S. (2018). Understanding auc-roc curve. *Towards Data Science*, *26*(1), 220–227.

Ragan, A. (2018). Taking the confusion out of confusion matrices. *Towards Data Science*.

Thanda, A. (2022). What is logistic regression? a beginner's guide.

Tracy, J., & Esposito, J. (2019). *Enlisted green platoon training in the special operations forces: Yesterday and today.* Retrieved 2023-04-04, from https://arsof-history.org

Vinnedge, A., Schwartz, B., Baller, D., & Dykhuis, E. (2022). Utilizing machine learning models to predict success in special operations assessment.

Winkie, D. (2022). Army special operations command aims to reverse recruiting woes.

Yiu, T. (2019). *Understanding random forest: How the algorithm works and why it is so effective.* Retrieved from https://towardsdatascience.com/understanding-random-forest-58381e0602d2

Zazanis, M. M. (1999). *Prescreening methods for special forces assessment and selection* (Vol. 1094). US Army Research Institute for the Behavioral and Social Sciences.